A major obstacle for the detection of low-mass exoplanets with the radial velocity (RV) method is the signal caused by stellar activity, which can overwhelm the planet signal. This paper presents a new method for dealing with the activity signal - the Skew Normal (SN) distribution as opposed to the Normal (Gaussian) distribution for fitting the CCF. The basis for this technique is that stellar activity in the form of spots, plage, etc. will alter the shapes of the spectral line profiles and this mimics the signal of a planet. This activity introduces an asymmetry in the CCD, which traditionally is measured via line bisectors. Since the SN distribution can model this one can, in principle, achieve an RV measurement that is less sensitive to activity.

The SN technique is interesting and the underlying concept seems promising. Unfortunately I do not think that the results of this paper merit publication. Believe me, it is not easy for me to outright reject a paper. I rarely do this. But in this case I cannot recommend publication of this paper based on three major criticisms:

1. The authors have failed to prove the advantages of using the SN approach. First, they demonstrate that the SN method on real data *consistently* produces an RV precision that is 60% larger than the Normal method. Why on earth would any researcher want to use a method that increases your rms scatter of your initial measurements by a factor of 1.6? Second, the authors promote the SN as a more efficient method for identifying and removing the activity signal from the RVs. However, when all is said and done, and the activity correction is applied to the initial RVs (with higher scatter!) one gets the same answer that one gets when applying the traditional Normal method. I could certainly understand tolerating a 60% increase in the RV standard deviation if it meant that after correcting for the activity signal you get a better measurement of the RV variations due to planets. But this seems to not be the case. The authors have simply not demonstrated a compelling reason why the community should use the SN method.
2. The authors make a bold claim that the SN method is more efficient at studying the underlying activity variations, but this has not really been demonstrated. This could have easily been done by isolating the activity RV variations from the measurements and seeing how the RVs from SN can better detect rotation periods, activity cycles, etc. In fact, the SN results show a higher RV scatter, but the authors never show if this scatter has structure (e.g. from rotation) or is purely random (inferior technique). In this respect I am quite surprised that the authors applied the SN to real stars where you do not know the answer, rather than simulated data where you do. Whenever you develop a new method the first step is to apply it to simulated data. Once you explore the advantages, limitations, and pitfalls of the method you then apply it to real data. Why hadn't this been done?

   I do not understand why the authors did not do this especially there is available software for generating this synthetic data, namely SOAP. Even more surprising, one of the co-authors is quite familiar with SOAP and has produced a wonderful data set of RVs from active stars that was used in the so-called "RV Challenge". This data set had known signals from activity and planets for a number of fake stars. Why didn't the authors apply the SN method to this data set first? I believe that SOAP delivers integrated spectral lines, so it should be easy to produce a CCF for testing with SN. You can then see how well SN recovers the known activity signal. This should be done first before applying the method to stars for which you do not have knowledge of the activity signal.

3. This paper is simply not well written. There are many instances where the authors were not clear, left out vital details, had poorly written expressions, and showed a lack of understanding for the details that they were discussing. An experienced writer in English should first read any revised version before submitting.

In conclusion, I do not think that this paper has results that are of particular interest to the exoplanet community. The authors have proposed a new method, but have failed to show that is of any real use. Some form of the paper might be published eventually. Maybe if the authors first applied it to simulated data, or outlined a clear path for improving the method and actually making it an effective tool. But in the current form I do not see the usefulness of the SN method.

Below are my detailed and numerous comments:

<u>Abstract:</u>

1) *"...these changes* [CCF*] are due to stellar activity*..."

You cannot generalize that all changes to the CCF are due to activity. Non-radial pulsations will also produce variations. It is true that this is not a concern in solar like stars, but you should state this explicitly. Also a blend by binary companion (SB1) with very low mass (i.e. low luminosity) will also produce bisector variations. So you need to specify that you are only concerned with RV and bisector variations due to activity.

2) *"...CCFs are more sensitive to activity which is a huge gain to probe stellar activity."*

"huge" is not a very quantitative description and it imparts no useful information to the reader. This is also counter to the results in this paper which shows that taking the Normal distribution and correcting for the RVs using BIS and FWHM quantities, you pretty much arrive at the same answer as the SN method. So where is the huge gain? Two sentences further in the abstract you even admit there is no advantage: *"However, once correcting the RVs from stellar activity signanol [SIC] using a linear combination of the CCD asymmetry and width, the RV residuals obtained from a Normal or SH fitting are very similar."*

3) "*RVs derived from the SN distribution are more sensitive to activity which is interesting.."*

" *interesting*" is a highly subjective and not very precise or quantitative expression. Interesting in what sense? I would term this a "sentence of no value", i.e. one that conveys not useful information to the reader.

4) "*Besides all these advantages, the errors on the RVs measured by SN method are 60% greater."*

To me this statement alone tells me that the SN method is not very useful, and not one that I would use. Any method that produces larger errors *before* correcting for the activity is not a very useful method. Why should the reader continue beyond this point?

5) "*Conclusion: We strongly encourage the use of the SN distribution to derive the different moments of the CCF, because the derived moments probe better stellar activity signals than when using a Normal distribution.*"

You simply cannot make this claim. What you have shown is that the SN method produces errors that are 60% larger than the Normal distribution and you hint that this is due to it being more sensitive to activity. You have not demonstrated this by looking at the "activity" signal that the SN isolates to see if it is actually more sensitive at finding rotation periods, activity cycles, etc. Where is the sensitivity in actually getting out useful information? Furthermore, after correcting the SN RVs for the activity, you arrive at the same answer as the Normal distribution method. I would conclude that the SN offers no real advantages over Normal so the community can continue to use the standard method.

Grammatically this is an ambiguous statement. Are you looking for "better signals", i.e. ones that you like, or probing stellar signals in a better way? (This is what you meant, but not what was stated.)

Introduction:

6) "*The radial velocity of a star is defined by the velocity of the center of mass of the star along the line of sight*. "

No, the radial velocity you measure of the star has many components that contribute to the RV: the space motion (present even if there is no companion), the motion about the center of mass (the so-called γ-velocity), oscillations, convective blue shift, etc.
You are just assuming that you are measuring the reflex motion of a star-planet system. There are many physical processes that produce a Doppler shift besides exoplanets.

7) "*This quantity can be derived precisely by measuring the Doppler shift of spectral lines..."*

Well, not if you have a poor measurement precision. You only get a "precise" measurement with simultaneous wavelength calibration. Again, the authors are not being very "precise" here.

8) In paragraph 1 you give two lists of spectrographs. I presume the first one is a list spectrographs that use the iodine method, although that is not explicitly expressed. The authors talk about the iodine cell and then just give a parenthetical list of spectrographs. There is no explicit statement that these spectrographs, designed for other purposes, are all equipped with an iodine cell. The reader must assume this. Furthermore, the list is not all-inclusive. I can think of a dozen facilities that use iodine cells. So you should say, "For example" or..."to name a few"

The second list it is for simultaneous Th-Ar calibration. Not all of these are "stabilized" like the authors claim. CORALIE, if I am not mistaken, was simultaneous Th-Ar without stabilization. That is why they built HARPS. Furthermore, all of theses spectrographs were designed for precise stellar RVs, unlike the spectrographs listed for the iodine method.

9) Paragraph 1: "*In this case if the spectrograph shifts are due to changing atmospheric conditions the iodine and stellar spectra are shifted in the same way*."

This is a poorly expressed thought that is technically wrong. The spectrograph does not shift due to changing atmospheric conditions (earth's atmosphere?). It changes because of mechanical shifts (vibrations, temperature changes in spectrograph housing, changes in instrumental profile) as well as movement of the image on the slit or fiber which yes, does affect stellar and iodine lines in the same way.

10) *"...one has to decorrelate the iodine spectrum from the stellar spectrum."*

No, you do not decorrelate (and I am not sure what the authors mean by this word) the iodine from the stellar spectrum. The authors show that they have no understanding how the method works. What is actually done is that a high-resolution iodine spectrum is combined with a spectrum of the star without iodine lines and a fit is made to the observed star+iodine spectrum. It is not decorrelation, but rather χ^2 fitting.

11) "*For spectrographs that are stabilised the spectrum of the calibration lamp is recorded close to the stellar spectrum on the CCD which prevents contamination of the stellar spectrum*."

This is not true! I have taken spectra with HARPS and there is cross-talk between spectra and Th-Ar particularly for strong emission lines. So there is some contamination. Furthermore, this "preventing of contamination" as the authors state has nothing to do with the spectrograph being stablized, it has to do with the simultaneous wavelength calibration using two fibers. In fact one reference the authors give, CORALIE, is not a stabilized spectrograph!

12) Paragraph 3. Regarding the C-shape of the bisector. The classic C-shape is only for late-type stars. In fact for hotter stars (F-type) it reverses.

The strength of the asymmetry depends not just on the velocities, it is more complicated than that. It also depends on the flux ratio between hot cells and cool lanes, as well as the ratio of surface areas between the two.

13) "*This small asymmetry modifies however slightly the estimated RVs of the star, reducing the accuracy of the measurement, but if this asymmetry does not vary with time, the precision is kept*."

But the activity causes the asymmetry to change due to rotation, spot evolution, etc. The authors are confusing the readers here. They talk about a changing activity signal, but now argue that the precision is only kept if the asymmetry is constant, which it is not. I think what they want to say is that you measure a precise stellar radial velocity, but the activity makes its own contribution which reduces the accuracy of the RV determination for the barycentric motion due to a planet.

14) "*The line asymmetry is commonly retrieved...or the bisector inverse slope span*"

It is the slope or span, not both. And you generally do not calculate the inverse, just the span. This may be correlated or inversely correlated with the RV.

15) Yes, for slowly rotating stars it is harder to measure the asymmetry, but then again it is a smaller effect. Rotation amplifies the effect.

16) "*All those parameters are correlated when stellar activity is dominant, and performing a step-by-step approach makes it difficult to correctly derive the errors on the different parameters retrieved*."

I simply do not understand this statement, possibly because it does not follow a standard English construction. You have a CCF and you measure a width and bisector span. This is a measurement error determined by the S/N, the resolution, and the stability of the spectrograph. You can determine an error in a quantity (RV and BIS). You can also have an uncertainty in the contribution of activity to the RV signal, but that is not a formal error. True errors come from photon statistics, systematic errors, instrumental errors, etc. You can attempt to remove the signal due to activity and that has an associated error, but that error is due to your lack of knowledge of the surface structure causing this RV and the exact value of the contribution to the RV. This is different from a measurement error. Maybe that is what you mean here, but it is not clear from what is written.

<u>Section 2 the SN distribution</u>

17) I cannot find where R, R+, x and w were defined. It is a common practice when you introduce variables in a paper, these should all be defined.

18) Paragraph 4: "*By using the so-called direct parameterization (DP) defined in Eq. 2 problems with parameters estimation arise in the evaluation of the likelihood function in a neighborhood of $\alpha = 0$, where the presence of a stationary point complicates the achievement of the maximum likelihood or least squares estimates*."

A poorly constructed sentence that simply confuses me! The authors are trying to say too many things in one sentence which loses the reader.

19) To be honest, I could not follow all the mathematics. I really would have to read the referenced material, but I had little time. Maybe the details can be put in an appendix?

<u>Section 4</u>

20) paragraph 1: "*To correct for stellar activity signals, it is common to consider a linear combination of the RVs with the BIS SPAN and the FWHM (or and SN FWHM in the SN case).*"

Is this really common? I am not aware of it having wide use in the community. If so, references would support this. The authors have to distinguish what their research group and collaborators do from what the rest of the community does.

21) Equation 9:
Is this expression just for the RV from activity alone, or the measured RV from the star (contribution of all phenomena)? If it is just activity, you should put a subscript otherwise the readers will be confused.

In the text you refer to the variable γ_1, but Eq. 9 has no subscript in γ. You give one equation for Normal and SN, but in the text you state these variables have to be replaced by such and such when using SN, etc. It will be less confusing if you simply have two equations, one for Normal and one for SN and with their respective variables.

<u>Section 5: Illustrative examples:</u>

22) Section 5.1 paragraph 2:
"*The SN fit is therefore significantly more sensitive to stellar activity This can be explained by the fact that because the SN includes an asymmetry parameter...*"

How can you be sure that the higher scatter in your measurement error is not simply due to the fact that you have an inferior measurement technique? This can be answered by isolating the RV contribution from activity (from the asymmetry). Is this random (noise) or is there structure (activity).

23) Section 5.1 paragraph 3:
"*We see that the RVs measured by the SN fitting show more variations than the RVs measured by the Normal fitting."*

This is not at all clear from the figure. Is there a way to quantify this?

24) If you are interested in using the RV signal due to activity to measure the rotation period, effects of activity, etc., why not just show the RVs due to the activity signal alone.?

25) Section 5.2 paragraph 1:
It is interesting to see that, in this case, the slope of 0.645 is significantly different from the one found for Alpha Centauri B in Fig. 3.

Why is this interesting? You would expect it to be given that this is a different star! If something is interesting you have to point out to the reader why.

"*Figure 9 shows the correlation between and BIS SPAN to understand the link*

Grammatically incorrect sentence: "Fig. 9 shows to understand"?

You need an alternate construction: "Figure 9 shows the correlation between BIS SPAN and ?? which helps us to understand...."

Note: there is a correlation between two quantities which is not expressed in the original text.

26) Section 5.2 para 2
So if I understand this correctly, SN gives higher scatter in the RV, but you make a bigger correction to the activity. In the end you arrive at the same answer as the Normal distribution and using Eq. 9 . So why use SN?

27) Section 5.3

If you are going to refer to the star as HD 10700 you should use the same name in the section heading.

"*This is probably because HD10700 is at a very low activity level, similar to the Sun at its minimum activity phase."*

And how exactly do you arrive at this conclusion? It is ok to hypothesize that the star has a low activity level. But to claim it has the same activity level as the sun at solar minimum is completely unqualified and should be supported by evidence. This is especially true since HD 10700 is a late F-type star and thus may have a different activity behavior from a G2 star.

"present a slightly more important rms"

And just what is a "more important" rms compared to a less important one? The rms is a measurement of the properties of the data often with respect to a fit.

28) Fig 5. The differences in the results of the two methods appear negligible to me

29) Section 5.4

Thus activity correction does not seem to be very efficient at understand the RV variation.

Poor English. How can a correction be "efficient at understand [SIC] a variation?"

Revised construction: "Thus the activity correction does not seem to be useful in helping us understand the RV variations."

30) "*This is probably because of the presence of planetary signals in the RV data (Mayor et al. 2011). Note however that the information on the orbital phase of the planets is not present in Mayor et al. (2011) and we cannot therefore remove those signals."*

So there is a planetary signal in the data but you do not have an orbit that gives you a phase? Then this is not a confirmed planetary signal and thus cannot support you claim

31) Section 5.5
"*The final star considered is Corot-7 which has low signal to noise*."

A star cannot have a "low signal to noise" only data has that. And what data are you referring to photometry (from CoRoT) RVs, from HARPS? You need to be clear.

paragraph 2:

"The RVs obtained with the SN show more variability than the RVs derived with the Normal. "

Hasn't this been the case for virtually all the stars you looked at?

32) Section 6, paragraph 2

"*we bootstrap a hundred times the stellar spectrum given..*"

It is not clear what you are doing here. I can understand bootstrapping a time series: you shuffle all the values keeping the time stamps fixed. How do you bootstrap a stellar spectrum? Do you shuffle the intensity values keeping the wavelengths fixed? That would produce a mess. Do you just add different levels of random noise? But the spectra already have noise in them, they are real observations. You need to state more clearly here what you did.

Section 6:

33) Paragraph 3
"However when fitting a SN to the CCF, we have RV errors that are on average 60% greater."

This seems to be the case in all your stars. I do not consider a method that delivers errors that are 60% larger to be a very effective method.

34) "*Thus the RVs measured by the SN do not have the same precision than RVs calculated using the Normal*."

So why use the SN method when it produces lower precision RV measurements?

Discussion:

35) "*Therefore, the SN fitting does not improve stellar activity correction when using a simple linear combination with the width and asymmetry of the CCF.*"

In this statement you have basically admitted that your method is not very useful. I do not see why anyone would be compelled to use it!

So, the bottom line is that you have introduced a method which produces similar results to the standard one. The only useful result of this paper is to convey to the readers that there is no significant advantage in implementing the SN method!